

Split plots in a Latin square: statistical models and formulas without and with subsampling

José Antonio Rodríguez-González¹

Delfina de Jesús Pérez-López²

Jesús Hernández-Ávila²

Artemio Balbuena-Melgarejo²

J. Ramón Pascual Franco-Martínez²

Andrés González-Huerta^{2,*}

1 Programa de Maestría en Ciencias Agropecuarias y Recursos Naturales-Facultad de Ciencias Agrícolas-Campus Universitario 'El Cerrillo'-Universidad Autónoma del Estado de México. El Cerrillo Piedras Blancas, Toluca de Lerdo, Estado de México. Tel. 722 2965552, ext. 117 (jrodriguezg008@alumno.uaemex.mx).

2 Centro de Investigación y Estudios Avanzados en Fitomejoramiento-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México. El Cerrillo Piedras Blancas, Toluca de Lerdo, Estado de México. AP. 435. Tel. 722 2965518, ext. 148 (djperetzl@uaemex.mx; jhernandeza@uaemex.mx; jrfrancom@uaemex.mx; abalbuenam@uaemex.mx; pcarn@uaemex.mx).

Autor para correspondencia: agonzalezh@uaemex.mx.

Abstract

Little information has been published for a split-plot arrangement for a Latin square design. This research builds statistical models and the formulas that allow obtaining degrees of freedom and sum of squares considering the methodologies of least squares and quadratic or matrix forms, without and with balanced subsampling. It is assumed that there is no interaction between rows, columns and factor A; however, it is indicated how to arrive at the same results by introducing the principle of crossing and nesting, particularly if InfoStat or InfoGen will be applied, it is suggested to use the main difference between the outputs generated by both statistical packages to subdivide the joint error into sampling error and experimental error; nevertheless, emphasis is also placed on the formulas that were derived for their direct calculation.

Keywords:

least squares, quadratic or matrix forms, two-factor experiments.



Introduction

Arrangements of experimental units in split plots (SP) or subsplit plots (SSP) have been widely used to study effects or variances in two- or three-factor trials (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016; Tirado and Tirado, 2017) or in series of experiments when they are extended to different environments, made up of years, localities, or combinations of both (González *et al.*, 2019; Pérez *et al.*, 2022).

The most commonly used randomization plans correspond to completely randomized experimental design (CRD) and randomized complete block design (RCBD) (Gomez and Gomez, 1984; Martínez, 1988; Little and Hills, 2008; Montgomery, 2009). Authors such as Ledolter (2010) discussed SP arrangements for complete trials and fractional factor experiments, with emphasis on CRD and RCBD, but without considering balanced subsampling.

The LSD, in SP or SSP, has not been widely used, but it could be very useful to evaluate the differences that originate from years, localities or both with distance between plants, fertilization formulas, organic fertilizers, insecticides, fungicides, herbicides, tillage methods, irrigation sheets, and forage cutting dates, among others, when these are assigned to the main plot and plant species or contrasting varieties of any of these could be considered in subplots (González *et al.*, 2019; Pérez *et al.*, 2022).

Farmers' land that is very heterogeneous also presents this type of random variability, which is undesirable to establish an adequate experiment; this could be controlled more efficiently by using a SP in LSD or another experimental design (González *et al.*, 2019; Pérez *et al.*, 2022).

Few reports have also been found on statistical models, analysis of variance, and comparison of means (Tirado and Tirado, 2017; González *et al.*, 2019), particularly when balanced subsampling and a statistical package are applied (Gomez and Gomez, 1984; Martínez, 1994; Montgomery, 2009) PROC Anova: Latin Square Split Plot:: SAS/STAT(R) 9.22 User's Guide.

Thus, the main objective of this research was to build their models without and with balanced subsampling and to present the formulas to calculate degrees of freedom and sum of squares based on two methodologies, as a prerequisite for the application of some statistical package.

Materials and methods

Symbology

In this study, the quantitative variable of interest will be identified with Y. Additionally, the symbology used will be that described by Mendenhall (1987), Sahagún (2007), and Montgomery (2009). The terminology described by Pérez *et al.* (2022) and González *et al.* (2023, 2004 a, b) will also be applied. The classification factors will be Rows, Columns, A, B, the levels of which will be, respectively: $i=1, 2, 3, h$; $j=1, 2, 3, c$; $k=1, 2, 3, \dots, t$; $l=1, 2, 3, \dots, b$. With balanced subsampling, in addition to the above, factor S ($m=1, 2, 3, s$) will be included.

Balanced statistical models

Relevant studies such as those by Tirado and Tirado (2017) presented a model for SP in LSD without subsampling, but the two presented below were constructed by applying the guide published by Sahagún (1998). Both models can also be built under the assumption of the absence of interaction between rows, columns, and levels of factor A, as suggested by Gomez and Gomez (1984); Martínez (1994); Montgomery (2009).

No subsampling

$$Y_{ijkl} = \mu + H_i + C_j + A_k + (HA)_{ik} + B_l + (AB)_{kl} + \varepsilon_{ijkl}$$

With subsampling

$Y_{ijklm} = \mu + H_i + C_j + A_k + (HA)_{ik} + B_l + (AB)_{kl} + \delta_{m(ikl)} + \epsilon_{ijklm}$. Where: Y is the variable of interest; μ is the overall arithmetic mean; H_i , C_j , A_k , and B_l are the effects caused by rows, columns, and factors A and B, respectively; $(HA)_{ik}$ is error a; $(AB)_{kl}$ is the interaction between both factors; $\delta_{m(ikl)}$ is the sampling error; ϵ_{ijkl} and ϵ_{ijklm} are the experimental error, without and with subsampling.

Experimental design

The levels of factor A will be assigned to main plots based on a Latin Square design, as suggested by Smith (1951); Martínez (1994); Tirado and Tirado (2017), and the levels of factor B are randomized into subplots in a completely random manner, one model will not consider balanced subsampling and the other will.

Recommended software

To obtain a randomization plan for this type of experiment, the Star statistical package can be used, which belongs to the International Rice Research Institute (IRRI), based in Los Baños, Philippines (Digital Tools | International Rice Research Institute (irri.org)). SAS (Statistical Analysis System; SAS OnDemand for Academics | SAS) or other statistical packages that generate them can also be employed.

To perform manual calculations with quadratic or matrix forms, the calculator freely available at <https://matrixcalc.org/es/> could be used. The data can be analyzed with the statistical packages referenced above or others, such as InfoGen and InfoStat (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016), the latter have been applied to divide the joint error into sampling and experimental errors in single-factor trials, for CRD, RCBD and LSD (González *et al.*, 2023).

Results

In main plots, rows (H), columns (C), and levels of factor A are equal ($h = c = t$); ht or ct is also equal to t^2 . For a RCBD, H or C, but not both, the number of replications chosen in the experiment could be considered as r .

Degrees of freedom (DF without subsampling)

DF total = $t^2b - 1$. DF H = $h - 1 = t - 1$; DF C = $c - 1 = t - 1$; DF A = $t - 1$; DF error a = $(t - 1)(t - 2)$. DF main plots (MP) = DF H + DF C + DF A + DF error a. For verification: DF MP = $t^2 - 1$. DF B = $b - 1$; DF AxB = $(t - 1)(b - 1)$. DF error b = $t(t - 1)(b - 1)$. DF subplots (SUB) = DF total - DF H - DF C - DF A - DF error a - DF B - DF AxB. Also: DF SUB = DF B + DF AxB + DF error b. For verification: DF SUB = $t^2(b - 1)$. DF MP + DF SUB = DF total = $(t^2 - 1) + t^2(b - 1) = t^2b - 1$.

Sum of squares (SS without subsampling)

In the denominator of the following formulas, h or c is null; H, C, A, and B will be represented by i , j , k , and l , respectively. Quadratic or matrix forms will be written as done by González *et al.* (2023); González *et al.* (2024 a, b), these will use sums or totals made on the subscript or subscripts that are not shown in their numerator.

$$SS \text{ total} = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2b} \right)^2 = Y'Y - \left(\frac{1}{t^2b} \right) Y'JY$$

$$SS H = \left(\frac{1}{tb} \right) \sum_{i=1}^h Y_{i...}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2b} \right)^2 = \left(\frac{1}{tb} \right) Y'_{i...} Y_{i...} - \left(\frac{1}{t^2b} \right) Y'JY$$

$$SS C = \left(\frac{1}{tb}\right) \sum_{j=1}^c Y_{.j.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2 b}\right)^2 = \left(\frac{1}{tb}\right) Y'_{.j.} Y_{.j.} - \left(\frac{1}{t^2 b}\right) Y' J Y$$

$$SS A = \left(\frac{1}{tb}\right) \sum_{k=1}^t Y_{.k.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2 b}\right)^2 = \left(\frac{1}{tb}\right) Y'_{.k.} Y_{.k.} - \left(\frac{1}{t^2 b}\right) Y' J Y$$

To calculate SS error a, SS TREAT1 must first be calculated, which is obtained as follows:

$$SS TRAAT1 = \left(\frac{1}{b}\right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2 b}\right)^2 = SS H + SS C + SS A + SS \text{ error a}$$

Also:

$$SS TREAT1 = \left(\frac{1}{b}\right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{t^2 b}\right) Y' J Y = SS \text{ main plots}$$

Therefore, SS error a = SS TREAT1 - SS H - SS C - SS A. To verify that SS error a is correct, apply the alternative formula:

$$SS \text{ error a} = \left(\frac{1}{b}\right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{1}{tb}\right) \sum_{i=1}^h Y_{i...}^2 - \left(\frac{1}{tb}\right) \sum_{j=1}^c Y_{.j.}^2 - \left(\frac{1}{tb}\right) \sum_{k=1}^t Y_{.k.}^2 + 2 FC =$$

$$\left(\frac{1}{b}\right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{tb}\right) Y'_{i...} Y_{i...} - \left(\frac{1}{tb}\right) Y'_{.j.} Y_{.j.} - \left(\frac{1}{tb}\right) Y'_{.k.} Y_{.k.} + \left(\frac{2}{t^2 b}\right) Y' J Y$$

From the above, it can also be verified that the sum of squares of the main plots (SS MP) is equal to:

$$SS MP = \left(\frac{1}{b}\right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - FC = \left(\frac{1}{b}\right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{t^2 b}\right) Y' J Y$$

$$SS B = \left(\frac{1}{t^2}\right) \sum_{l=1}^b Y_{...l.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2 b}\right)^2 = \left(\frac{1}{t^2}\right) Y'_{...l.} Y_{...l.} - \left(\frac{1}{t^2 b}\right) Y' J Y$$

To calculate SS Ax B, SS TREAT2 must first be obtained as follows:

$$SS TREAT2 = \left(\frac{1}{t}\right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}}{t^2 b}\right)^2 = \left(\frac{1}{t}\right) Y'_{.kl.} Y_{.kl.} - \left(\frac{1}{t^2 b}\right) Y' J Y$$

The denominator of the first part of the formula must be h or c, but not both and since h = c = t, that is why t, the number of levels in factor A, was written. SS TREAT2 = SS A + SS B + SS Ax B. Therefore: SS Ax B = SS TREAT2 - SS A - SS B.

$$SC \text{ Error } b = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}^2 - \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{ik.}^2 - \left(\frac{1}{t} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl}^2 + \left(\frac{1}{tb} \right) \sum_{k=1}^t Y_{.k.}^2$$

$$= Y'Y - \left(\frac{1}{b} \right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{t} \right) Y'_{.kl} Y_{.kl} - + \left(\frac{1}{tb} \right) Y'_{.k.} Y_{.k.}$$

$$SC \text{ SUB} = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b Y_{ijkl}^2 - \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{ik.}^2 = Y'Y - \left(\frac{1}{b} \right) Y'_{i.k.} Y_{i.k.}$$

Degrees of freedom (DF) with subsampling

Due to the addition of factor S (m= 1, 2, 3, s): DF total= $t^2bs - 1$. DF H= $h - 1 = t - 1$. DF C= $c - 1 = t - 1$. DF A= $t - 1$. DF error a= $(t - 1)(t - 2)$. DF MP= DF H + DF C + DF A + DF error a.

Also: DF MP= $t^2 - 1$. DF B= $b - 1$. DF AxB= $(t - 1)(b - 1)$. DF JE= DF total - DF H - DF C - DF A - DF error a - DF B - DF AxB. Where: DF JE are the degrees of freedom of the joint error. For its part: DF JE= $t(tbs - t - b + 1)$. In addition, it is known that (González *et al.*, 2023): DF JE= DF SE + DF EE.

But DF EE= $t(t - 1)(b - 1)$, therefore: DF SE= DF JE - DF EE= $t^2b(s - 1)$. For verification: DF SUB= $t^2(bs - 1)$. DF MP + DF SUB= $(t^2 - 1) + t^2(bs - 1) = t^2bs - 1 = \text{DF total}$.

Sum of squares (SS) with subsampling

$$SS \text{ total} = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = Y'Y - \left(\frac{1}{t^2bs} \right) Y'JY$$

$$SS \text{ H} = \left(\frac{1}{tbs} \right) \sum_{i=1}^h Y_{i...}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{tbs} \right) Y'_{i..} Y_{i..} - \left(\frac{1}{t^2bs} \right) Y'JY$$

$$SS \text{ C} = \left(\frac{1}{tbs} \right) \sum_{j=1}^c Y_{.j...}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{tbs} \right) Y'_{.j..} Y_{.j..} - \left(\frac{1}{t^2bs} \right) Y'JY$$

$$SC \text{ A} = \left(\frac{1}{tbs} \right) \sum_{k=1}^t Y_{..k..}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{tbs} \right) Y'_{..k.} Y_{..k.} - \left(\frac{1}{t^2bs} \right) Y'JY$$

SS MP= SS H + SS C + SS A + SS HxA. In this one, SS HxA= SS error a. Thus:

$$SS \text{ MP} = SC \text{ TREAT1} = \left(\frac{1}{bs} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{bs} \right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{t^2bs} \right) Y'JY$$

Therefore: SS HxA= SS TREAT 1-SS H-SS C-SS A=

$$\left(\frac{1}{bs} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{1}{tbs} \right) \sum_{i=1}^h Y_{i...}^2 - \left(\frac{1}{tbs} \right) \sum_{j=1}^c Y_{.j...}^2 - \left(\frac{1}{tbs} \right) \sum_{k=1}^t Y_{..k..}^2 + 2 \left(\frac{\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2$$

Also:

$$SS H \times A = \left(\frac{1}{bs} \right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{tbs} \right) Y'_{i...} Y_{i...} - \left(\frac{1}{tbs} \right) Y'_{.j...} Y_{.j...} - \left(\frac{1}{tbs} \right) Y'_{..k.} Y_{..k.} + \left(\frac{2}{t^2bs} \right) Y'Y$$

$$SS B = \left(\frac{1}{t^2s} \right) \sum_{l=1}^b Y_{...l}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^b \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{t^2s} \right) Y'_{...l} Y_{...l} - \left(\frac{1}{t^2bs} \right) Y'Y$$

To calculate SS AxB, SS TREAT2 must first be calculated as follows:

$$SC TREAT2 = \left(\frac{1}{ts} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl}^2 - \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^b \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 = \left(\frac{1}{ts} \right) Y'_{.kl} Y_{.kl} - \left(\frac{1}{t^2bs} \right) Y'Y$$

$$SS A \times B = SS TREAT2 - SS A - SS B =$$

$$\left(\frac{1}{ts} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl}^2 - \left(\frac{1}{tbs} \right) \sum_{k=1}^t Y_{.k.}^2 - \left(\frac{1}{t^2s} \right) \sum_{l=1}^b Y_{...l}^2 + \left(\frac{\sum_{i=1}^h \sum_{j=1}^t \sum_{k=1}^b \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}}{t^2bs} \right)^2 =$$

$$\left(\frac{1}{ts} \right) Y'_{.kl} Y_{.kl} - \left(\frac{1}{tbs} \right) Y'_{.k.} Y_{.k.} - \left(\frac{1}{t^2s} \right) Y'_{...l} Y_{...l} + \left(\frac{1}{t^2bs} \right) Y'Y$$

Now, it is possible to obtain, by difference, the SS JE, which is the sum of squares of the joint error. Its value is estimated from: SS JE= SS total-SS H - SS C-SS A-SS error a-SS B-SS AxB.

It is also verifiable as follows:

$$SS JE = \sum_{i=1}^h \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}^2 - \left(\frac{1}{bs} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{1}{ts} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl}^2 + \left(\frac{1}{tbs} \right)$$

$$\sum_{k=1}^t Y_{.k.}^2 = Y'Y - \left(\frac{1}{bs} \right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{ts} \right) Y'_{.kl} Y_{.kl} + \left(\frac{1}{tbs} \right) Y'_{.k.} Y_{.k.}$$

$$SS \text{ sampling error (SE)} =$$

$$\sum_{i=1}^h \sum_{k=1}^t \sum_{l=1}^b \sum_{m=1}^s Y_{ijklm}^2 - \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t \sum_{l=1}^b Y_{i.kl}^2 = Y'Y - \left(\frac{1}{b} \right) Y'_{i.kl} Y_{i.kl}$$

Additionally, the SS of experimental error (SS EE) is equal to:

$$SS EE = \left(\frac{1}{b} \right) \sum_{i=1}^h \sum_{k=1}^t \sum_{l=1}^b Y_{i.kl}^2 - \left(\frac{1}{bs} \right) \sum_{i=1}^h \sum_{k=1}^t Y_{i.k.}^2 - \left(\frac{1}{ts} \right) \sum_{k=1}^t \sum_{l=1}^b Y_{.kl}^2 + \left(\frac{1}{tbs} \right) \sum_{k=1}^t Y_{.k.}^2$$

$$= \left(\frac{1}{b} \right) Y'_{i.kl} Y_{i.kl} - \left(\frac{1}{bs} \right) Y'_{i.k.} Y_{i.k.} - \left(\frac{1}{ts} \right) Y'_{.kl} Y_{.kl} + \left(\frac{1}{tbs} \right) Y'_{.k.} Y_{.k.}$$

Discussion

Arrangements in SP and subsplit plots for CRD and RCBD have been used more frequently in trials conducted in agricultural and forestry sciences, among others (Gomez and Gomez, 1984; Martínez, 1988, 1994; Montgomery, 2009; González *et al.*, 2019), but the corresponding to the Latino square (LSD) is not well documented (Sahagún, 1998; Ledolter, 2010; Tirado and Tirado, 2017; González *et al.*, 2022).

In this study, LSD was applied to the levels of factor A in main plots and those corresponding to factor B are assigned to the subplots completely at random, as suggested by Smith (1951); Martínez (1994); Tirado and Tirado (2017). This gives rise to two types of error: one equal to that of an LSD for a one-factor experiment and another that is the residual in the model (Martínez, 1994; <https://biometrics.ilri.org/Publication/Full%20Text/chapter20.pdf>).

In the published literature, little evidence has been found in relation to the statistical reference model, as well as for an Analysis of Variance (Anova) and a comparison of means (Ledolter, 2010; Tirado and Tirado, 2017; <https://biometrics.ilri.org/Publication/Full%20Text/chapter20.pdf>), especially for subsampling (Gomez and Gomez, 1984; Martínez, 1988; proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide).

In this sense, Martínez (1994) only presented a table with the sources of variation (SV) and the degrees of freedom (DF) that could be calculated without subsampling. Gomez and Gomez (1984) showed two tables with the SV and DF for a RCBD and for a SP arrangement with balanced subsampling, but for a single trial; additionally, they described the procedures to obtain an Anova in both situations, with emphasis on the estimation of the sampling (ME) and experimental (EE) errors, which, in the present study, make up the joint error (JE).

In the statistical analysis system (SAS), only the data provided by Smith (1951; proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide), but without subsampling, and a code to obtain an Anova with three types of errors were presented; in relation to the present research, CxHxA is equal to error a and the sum of HxB and the residual of the model give rise to error b. Tirado and Tirado (2017) presented the statistical model for a SP in LSD without subsampling, as well as an example to generate their Anova; they also generalized this type of analysis for an arrangement of split plots in a LSD without subsampling.

This study built the statistical model for SP experiments in LSD, without and with balanced subsampling, and generated the formulas for calculating DF and sum of squares (SS) by applying least squares and quadratic or matrix forms, based on the recommendations made for other studies by González *et al.* (2023); González *et al.* (2024 a, b); however, they also highlighted the use of software.

This research considered that there is no interaction between rows (H), columns (C), and levels of factor A, as suggested by Gomez and Gomez (1984); Martínez (1988, 1994); Montgomery (2009); Tirado and Tirado (2017), among others, but applying the guide published by Sahagún (1998), the same results will be reached with the following considerations: H nested in C or vice versa; H, C, or both nested within factor A; factors A and B are crossed, H, C, or both nested in the AxB interaction; the joint error nested in all components of this model.

With balanced subsampling, in addition to the above, S will be nested in H, C, and AxB. When applying InfoGen or InfoStat, it will be correct to choose one of the following combinations as Error a: HxC, CxH, HxA, CxA, or HxCxA, because in an LSD, it is also true that $H=C=A=R$, where R is the number of replications if a CRD or a RCBD were chosen. In the present study, the HxA interaction was considered as error a, but when considering the code for SAS that allowed the analysis of Smith's (1951) data, it is equivalent to the CxHxA interaction.

Thus, the remaining components in the non-subsampled model will be B, AxB, and error b. With balanced subsampling, the JE should be divided into sampling (ME) and experimental (EE) errors and the remaining components will be the same as for the SP in LSD without subsampling. In both statistical packages, SE will be estimated directly, when the H*A*B>S instruction is captured in the specifications to the terms of the model, but C*A*B>S will produce the same results when applying InfoStat or InfoGen (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016).

The difference that arises when considering or not S, the number of times that balanced subsampling is applied will allow the indirect calculation of SS SE, but to validate results, it is possible to use the formulas previously presented, as well as those that were built to estimate JE and EE, if the user uses least squares, quadratic or matrix formulas, or both.

The STAR statistical package generates a randomization plan for SP in LSD, as well as an Anova and a comparison of means for factors A and B, and for their interaction, but the subsampling modality is not implemented in this package. This same situation has been observed when reviewing various statistical packages that are frequently used in agricultural and forestry sciences, among others (González *et al.*, 2019; Pérez *et al.*, 2022).

A table could also be constructed for the HxAxB interaction, the difference between total SS and SS of this triple interaction will produce SS SE. Thus, SS EE would be the difference between SS JE and SS SE. For a SP in LSD without subsampling, the statistical hypotheses related to H, C, and A will be tested using error a or the HxA interaction, whereas those corresponding to factor B and the AxB interaction will be evaluated with the joint error (Ledolter, 2010; Tirado and Tirado, 2017; Proc Anova: Latin Square Split Plot :: SAS/STAT(R) 9.22 User's Guide).

For the model with subsampling, it must first be determined whether SE is significant: if it is, it will be used to test the statistical hypotheses of factor B and its interaction; if it is not, both sources of variation will be evaluated using the residual of the model (Ledolter, 2010; Tirado and Tirado, 2017).

This same approach will be used to perform the mean comparison tests for the components within main plots and within subplots (Ledolter, 2010; Tirado and Tirado, 2017), but one can also resort to the recommendations provided by Gomez and Gomez (1984); Little and Hills (2008); Sahagún (1998), if H, C, but not both, were not significant, in which case such comparisons will be equivalent to those of a RCBD in SP.

If H and C are not significant either, the user has the option of analyzing their data as a CRD in SP using the same database as for the previously cited cases (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini and Di Rienzo, 2016; Tirado and Tirado, 2017).

The methodologies presented in this research will allow to standardize formulas to calculate degrees of freedom and sum of squares in an easy and reliable way, as shown by studies conducted by González *et al.* (2023); González *et al.* (2024 a, b), with their symbology and that described in Mendenhall (1987); Sahagún (2007); Montgomery (2009).

The quadratic or matrix forms that will most frequently feed matrix calculators or statistical packages, such as SAS, Agrobases, SPSS, StatGraphics, STAR, PB Tools, among others, will be generated easily and reliably, so the procedures to build a statistical model, to obtain an Anova, and to compare means of treatments in factorial experiments, for the type of arrangement of experimental units considered here, will allow the calculations relevant to both methodologies to be verified, directly or indirectly, faster and more reliably.

Conclusions

In the two statistical models that were built in this research, it was assumed that there is no interaction between rows, columns, and levels of factor A, assignable to main plots in a Latin square design; additionally, it was considered that there is crossover between factors A and B. Both models can also be generated if the following is considered: H nested in C or vice versa; H, C, or both nested within factor A; factors A and B are crossed; H, C, or both are nested in the AxB interaction; the joint error is nested in all components of this model.

With balanced subsampling, in addition to the above, S will be nested in H, C, and AxB. To verify degrees of freedom and sums of squares in both models, the alternative formulas that were constructed for the a, set, sample, and experimental errors can be applied for each or both methodologies. Another option would be to directly apply the formulas generated for the main plots and subplots and indirectly estimate the four reference errors.

Bibliography

- 1 Balzarini, M. G. y Rienzo, J. A. 2016. InfoGen. FCA. Universidad Nacional de Córdoba, Argentina. <http://www.info-Gen.com.ar>.
- 2 Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Rienzo, J. A. y Robledo, C. W. 2008. Manual del usuario de InfoStat. Ed. Brujas, Córdoba, Argentina. 348 p.
- 3 Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. y Robledo, C. W. 2008. InfoStat, versión 2008. Grupo InfoStat, FCA. Universidad Nacional de Córdoba, Argentina. <https://www.infostat.com.ar>.
- 4 Gomez, K. A. y Gomez, A. A. 1984. Statistical procedures for agricultural research. 2nd Ed. John Wiley & Sons, Inc. Printed in Singapore. 680 p.
- 5 González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F.; Franco, M. J. R. P. y Padilla, L. A. 2019. InfoStat, InfoGen y SAS para contrastes mutuamente ortogonales en experimentos en bloques completos al azar en parcelas subdivididas. Revista Mexicana de Ciencias Agrícolas. 10(6):1417-1431.
- 6 González, H. A.; Pérez, L. D. J.; Balbuena, M. A.; Franco, M. J. R.; Gutiérrez, R. F. y Rodríguez, G. J. A. 2023. Submuestreo balanceado en experimentos monofactoriales usando InfoStat y InfoGen: validación con SAS. Revista Mexicana de Ciencias Agrícolas. 14(2):235-249. Doi: <https://doi.org/10.29312/remexca.v14i2.3418>.
- 7 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R. P.; Rubí, A. M. y Balbuena, M. A. 2024 a. Tratamientos anidados dentro de un arreglo en grupos de bloques completos balanceados. Revista Mexicana de Ciencias Agrícolas. 15(2):3634. Doi.org/ 10.29312/remexca.v15i2.3634.
- 8 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R. P.; Balbuena, M. A.; Rubí, A. M. 2024b. Serie de experimentos para tratamientos anidados en grupos en arreglo de bloques completos balanceados. Revista Mexicana de Ciencias Agrícolas. 15(7):e3831. <https://doi.org/10.29312/remexca.v15i7.3831>.
- 9 Ledolter, J. 2010. Split-plot design: discussion and examples. International Journal of Quality Engineering and Technology. 1(4):441-457. Doi.org/ 10.1504/IJQET.2010.035588.
- 10 Little, T. M. , Hills F. J. 2008. Métodos estadísticos para la investigación en la agricultura. Ed. Trillas, SA de CV. México, DF. ISBN:978-968-24-3629-1. 270 p.
- 11 Martínez, G. A. 1988. Diseños experimentales. métodos y elementos de teoría. Editorial. Trillas, 1^a Ed. México, DF. 756 p.
- 12 Martínez, G. A. 1994. Experimentación agrícola: métodos estadísticos. Primera Ed. Universidad Autónoma Chapingo (UACH), México, DF. ISBN: 9789688842867. 357 p.
- 13 Mendenhall, W. 1987. Introducción a la probabilidad y la estadística. Grupo editorial Iberoamérica. 1^a Ed. México, DF. 626 p.
- 14 Montgomery, D. C. 2009. Design and analysis of experiments. 7thEd. John Wiley & Sons. Inc. U.S.A. 656 p.
- 15 Pérez, L. D.; Jasso, B. G.; Saavedra, G. C.; Franco, M. J. R. P.; Ramírez, D. J. F.; González, H. A. 2022. Uso de artificios en Opstat para analizar series de experimentos en dialélico parcial. Revista Mexicana de Ciencias Agrícolas. 13(2):273-287. Doi.org/ 10.29312/remexca.v13i2.3130.

- 16 Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos, aleatorios y mixtos. Departamento de Fitotecnia. Programa nacional de investigación en olericultura. Universidad Autónoma Chapingo (UACH). Boletín técnico núm. 2. 64 p.
- 17 Sahagún, C. J. 2007. Estadística descriptiva y probabilidad: una perspectiva biológica. 2ª Ed. Universidad Autónoma Chapingo (UACH). México, DF. ISBN: 97896802- 03574. 282 p.
- 18 Smith, W. G. 1951. Dissertation notes on Canadian sugar factories Ltd, Taber, Alberta, Canadá.
- 19 Tirado, E. G. y Tirado, G. D. N. 2017. Tratado de estadística experimental. primera edición. Ed centro de estudios e investigaciones para el desarrollo docente (CENID, AC), Guadalajara, Jalisco, México. ISBN: 978-607-8435-43-2. 202-264 pp.



Split plots in a Latin square: statistical models and formulas without and with subsampling

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 December 2024
Date accepted: 01 February 2025
Publication date: 15 April 2025
Publication date: Feb-Mar 2025
Volume: 16
Issue: 2
Electronic Location Identifier: e3926
DOI: 10.29312/remexca.v16i2.3926

Categories

Subject: Artículo

Keywords:

Keywords:

least squares
quadratic or matrix forms
two-factor experiments

Counts

Figures: 0
Tables: 0
Equations: 25
References: 19
Pages: 0